



Repository Interoperability

Interoperability and Preservation with the Hub and Spoke (HandS)

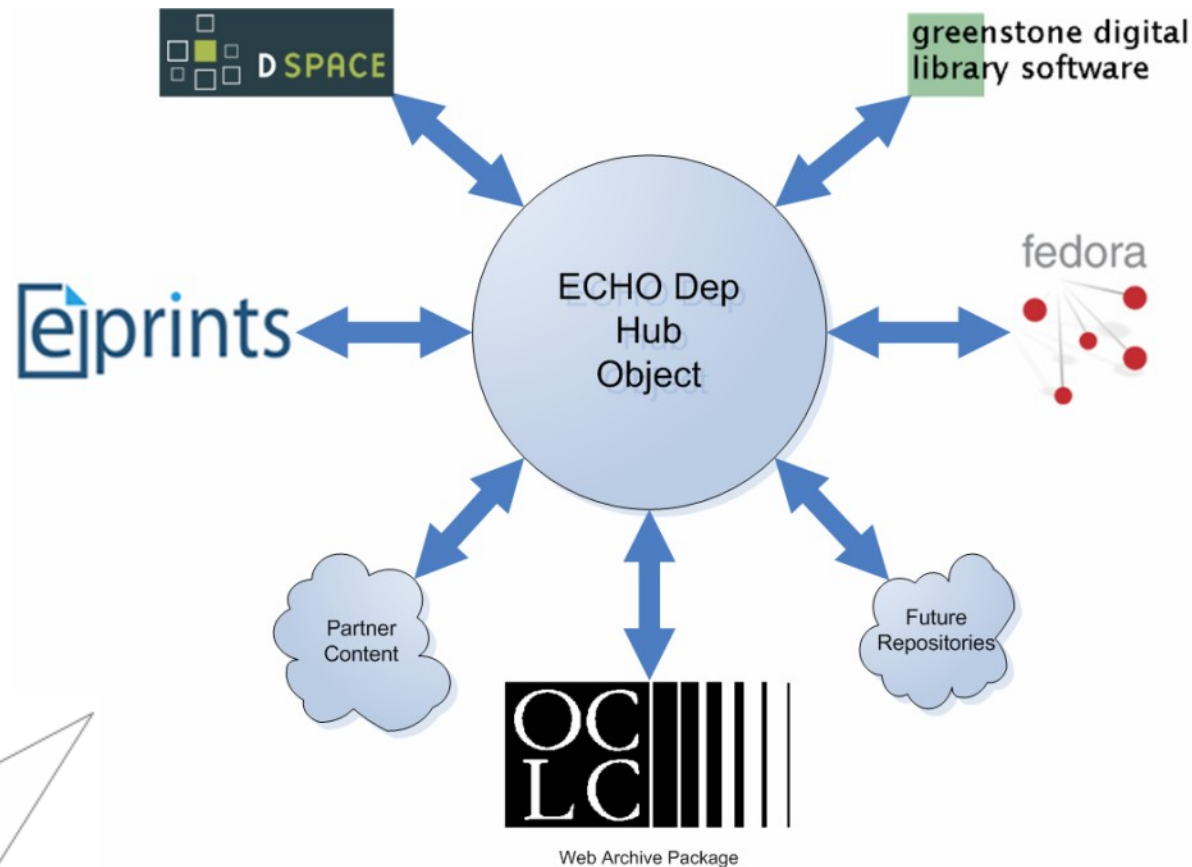


Tom Habing, Bill Ingram, Robert Manaster
University of Illinois Urbana-Champaign
thabing;wingram2;manaster@uiuc.edu



Key Ideas

- Interoperability is useful in its own right, but it is also important for preservation
 - Out-of-the-box repository interoperability is low
 - Institutions commonly rely on multiple repositories
 - Repository support for emerging preservation standards is low
 - Repositories change over time



For N different repositories that need to interoperate, this model reduces the complexity from $N*(N-1)$ to $2*N$.



This simple idea is the rationale for many different standards that aim to promote interoperability.

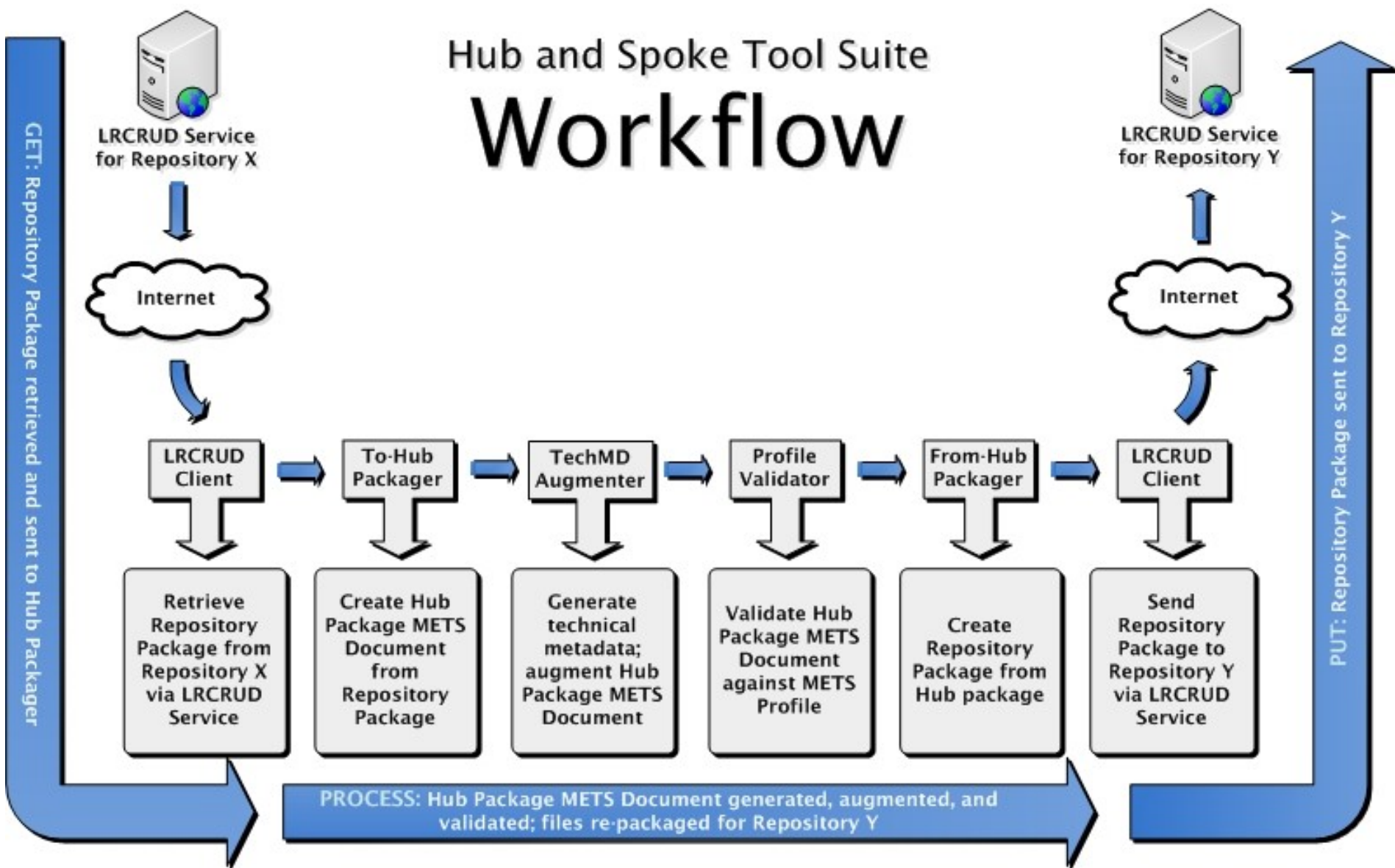


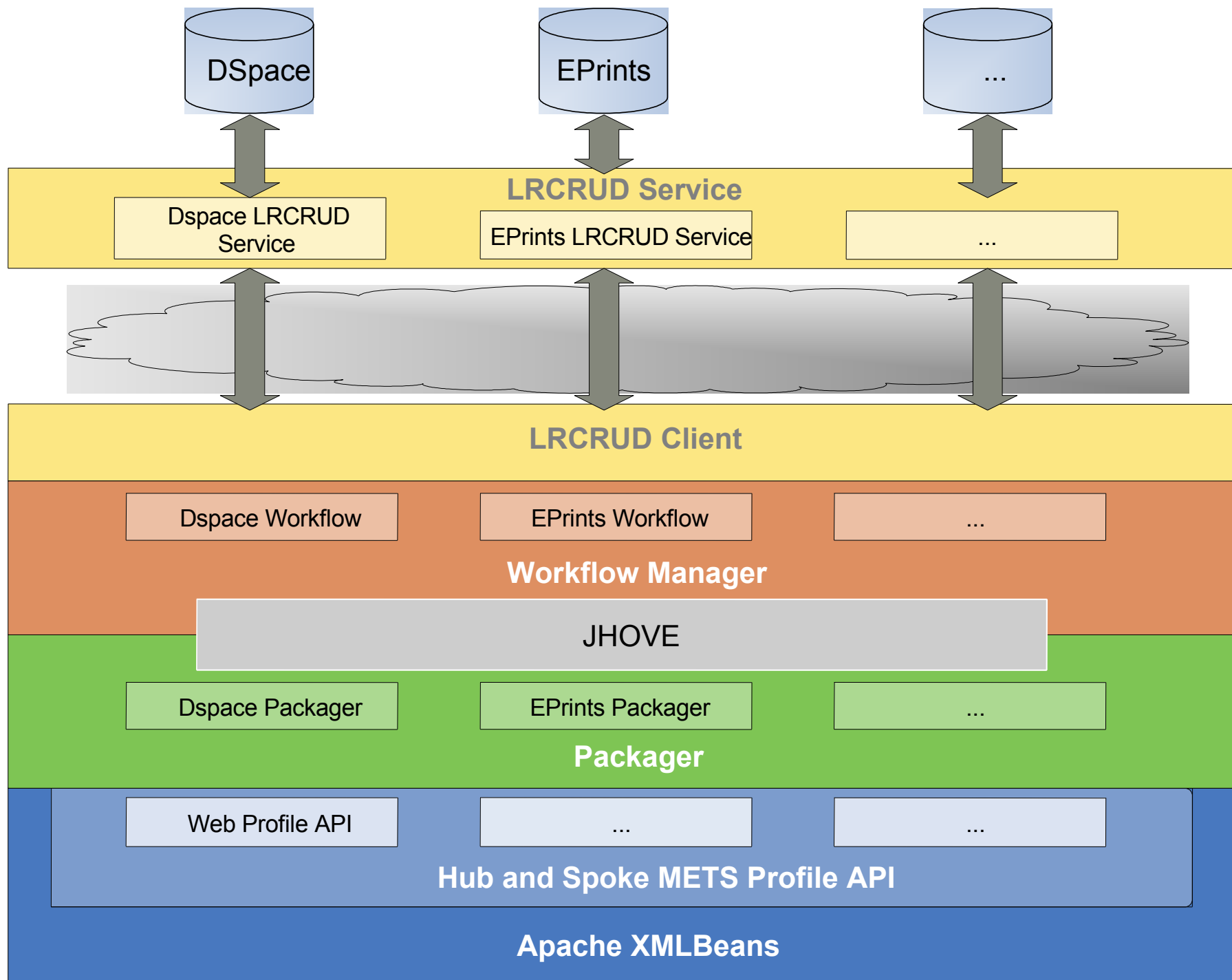
The Essentials

- Extensible METS profile
 - With versioning
- Repository specific processing and transformation utilities
- Java API for local integration and extensibility
 - Apache XML Beans
- Dissemination/Submission Web-service
 - RESTful



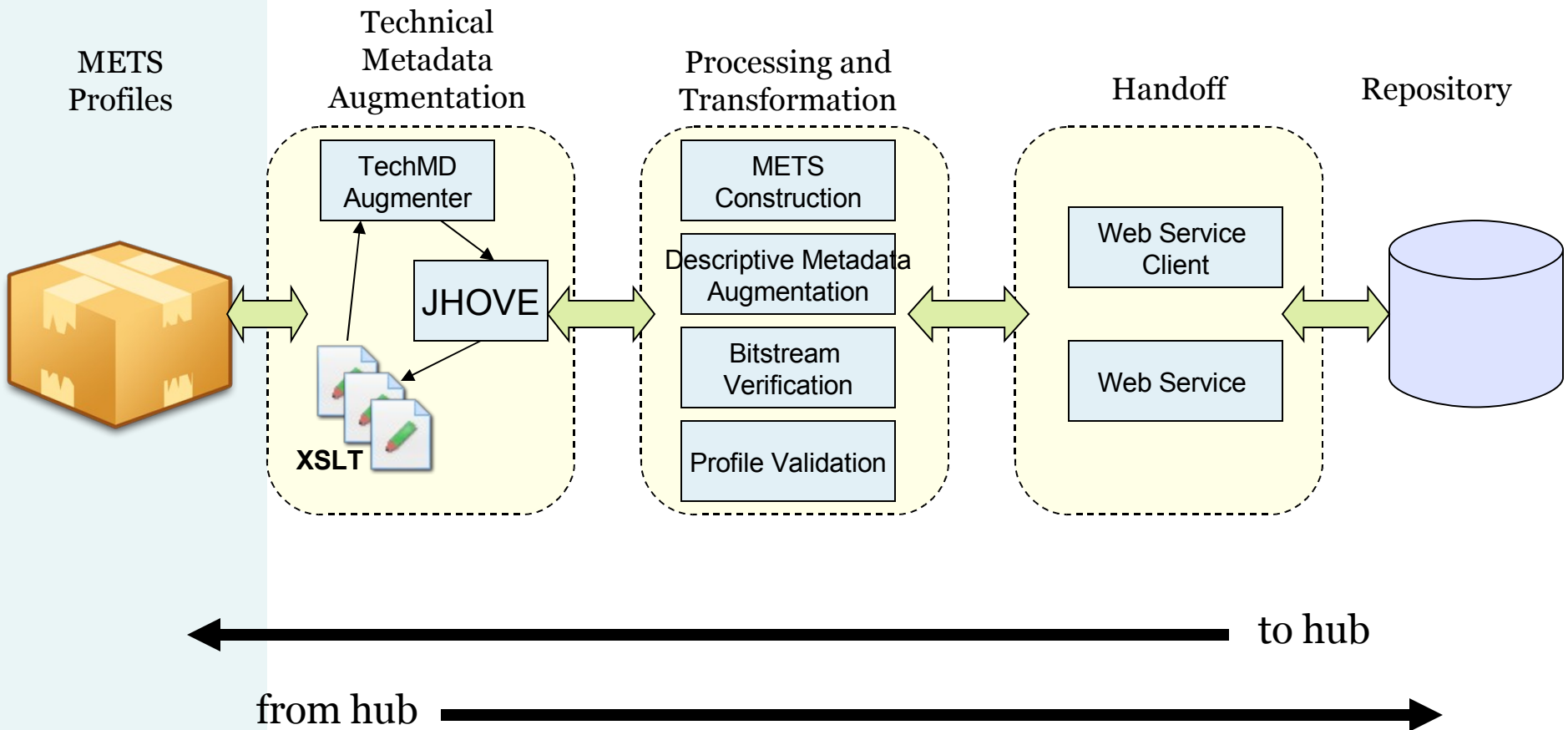
Hub and Spoke Tool Suite Workflow







Functional Overview



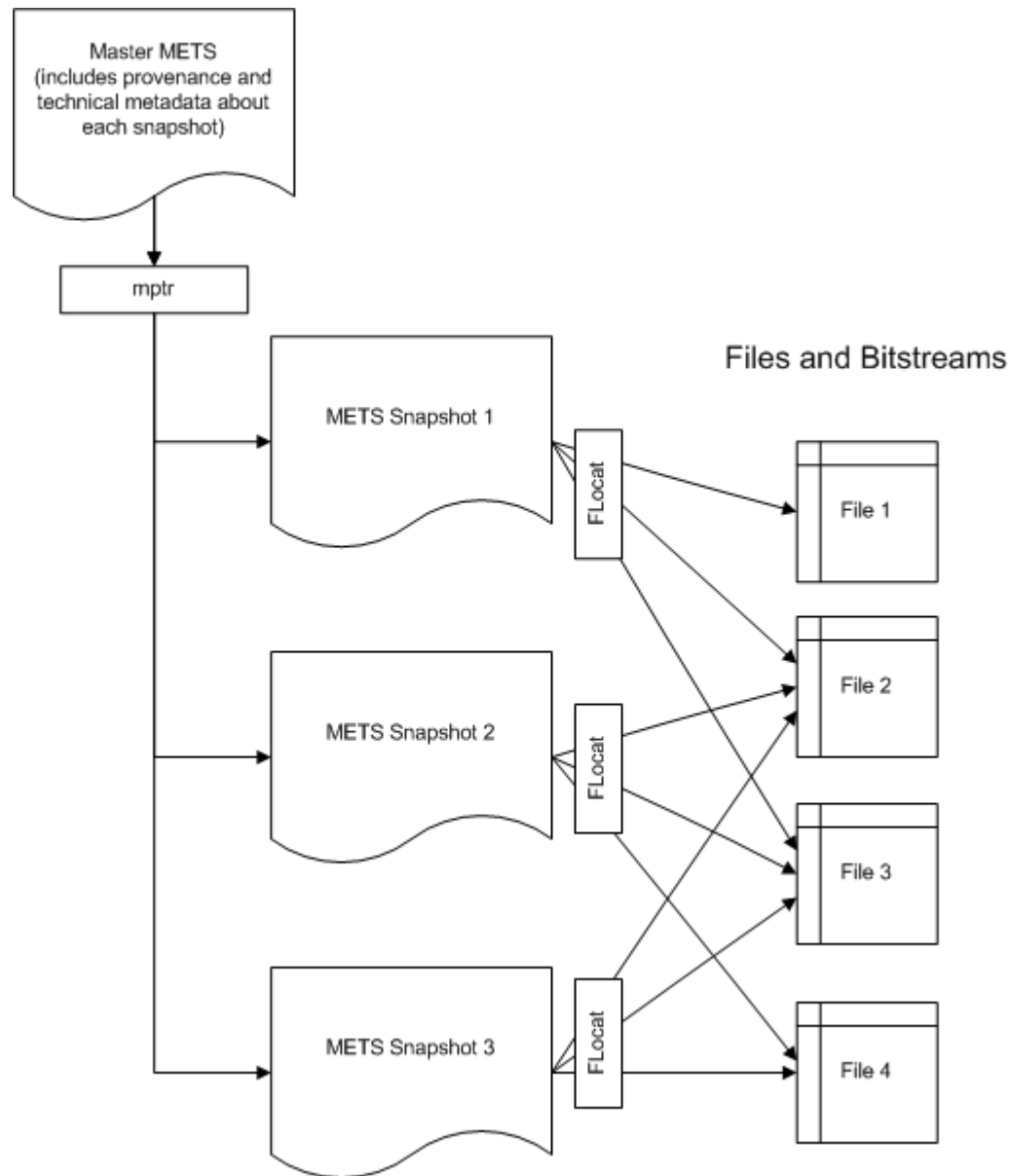


METS Profiles

- Non-prescriptive in regards to structure or file formats
- Intended to overlay other profiles which specify case-specific needs (i.e. web captures)
- PREMIS
- MODS
 - Must conform to the DLF Aquifer profile
- File-format specific technical metadata
 - MIX, VIDEO MD, AUDIO MD, others as appropriate

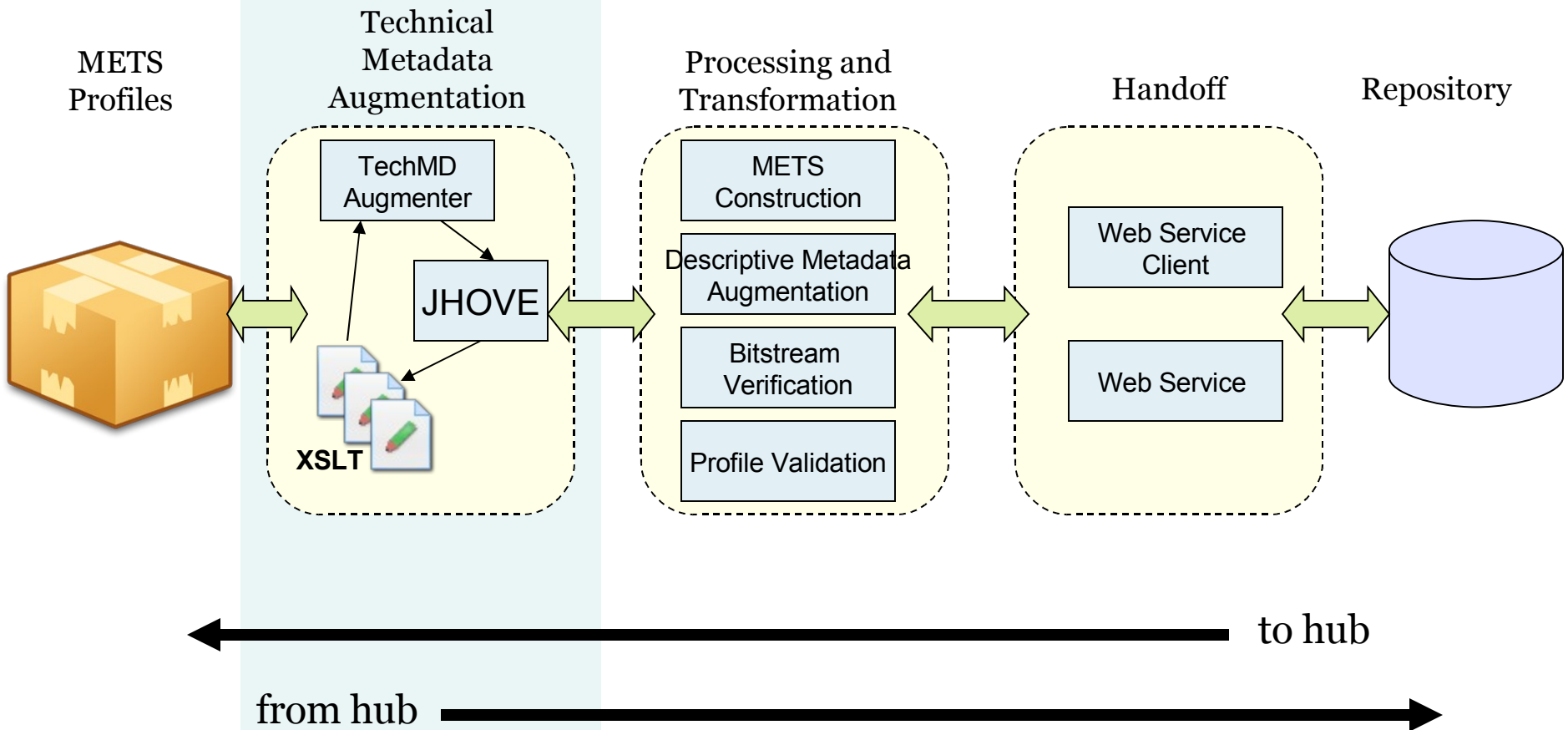


Master METS + Snapshots





Functional Overview



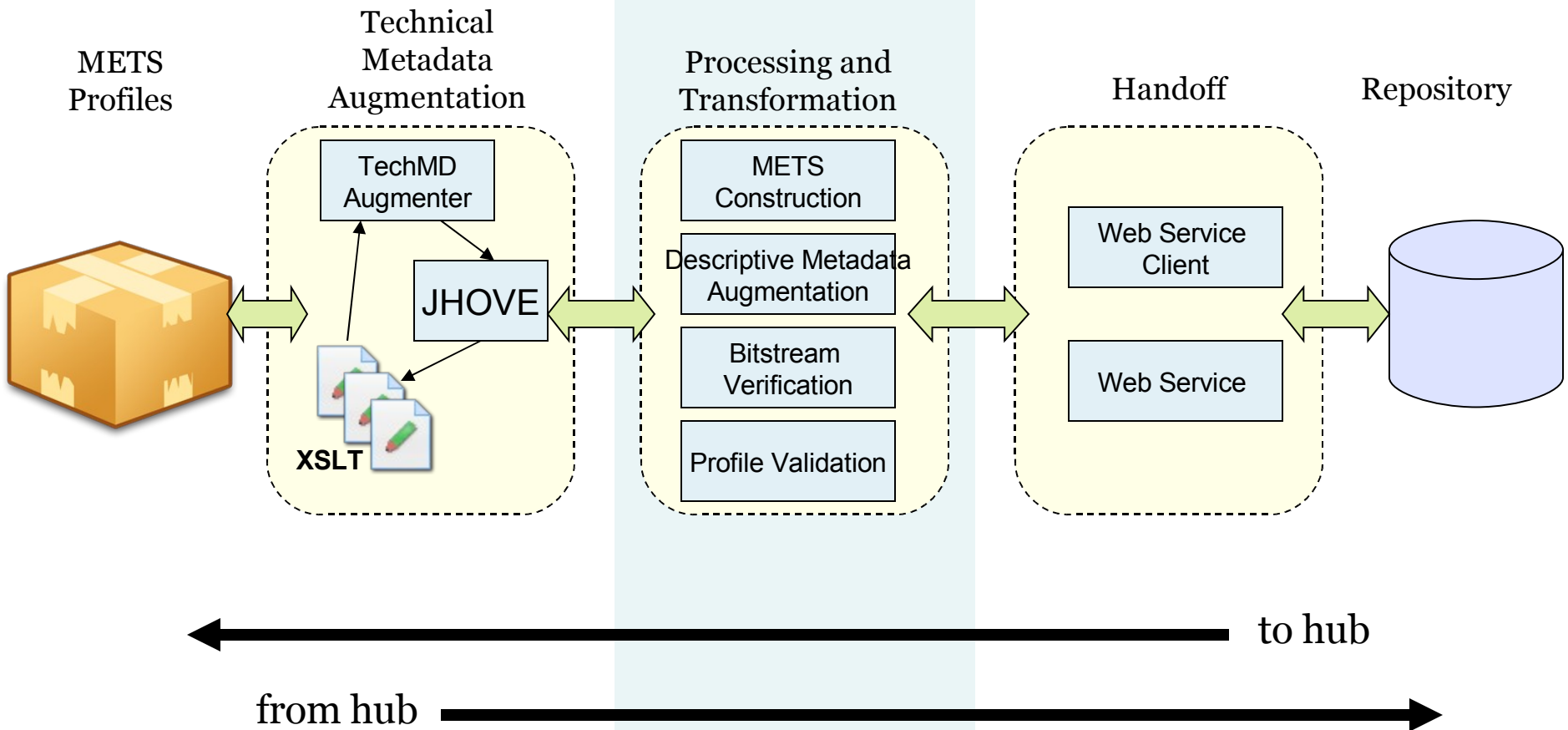


Technical Metadata Generation/Augmentation

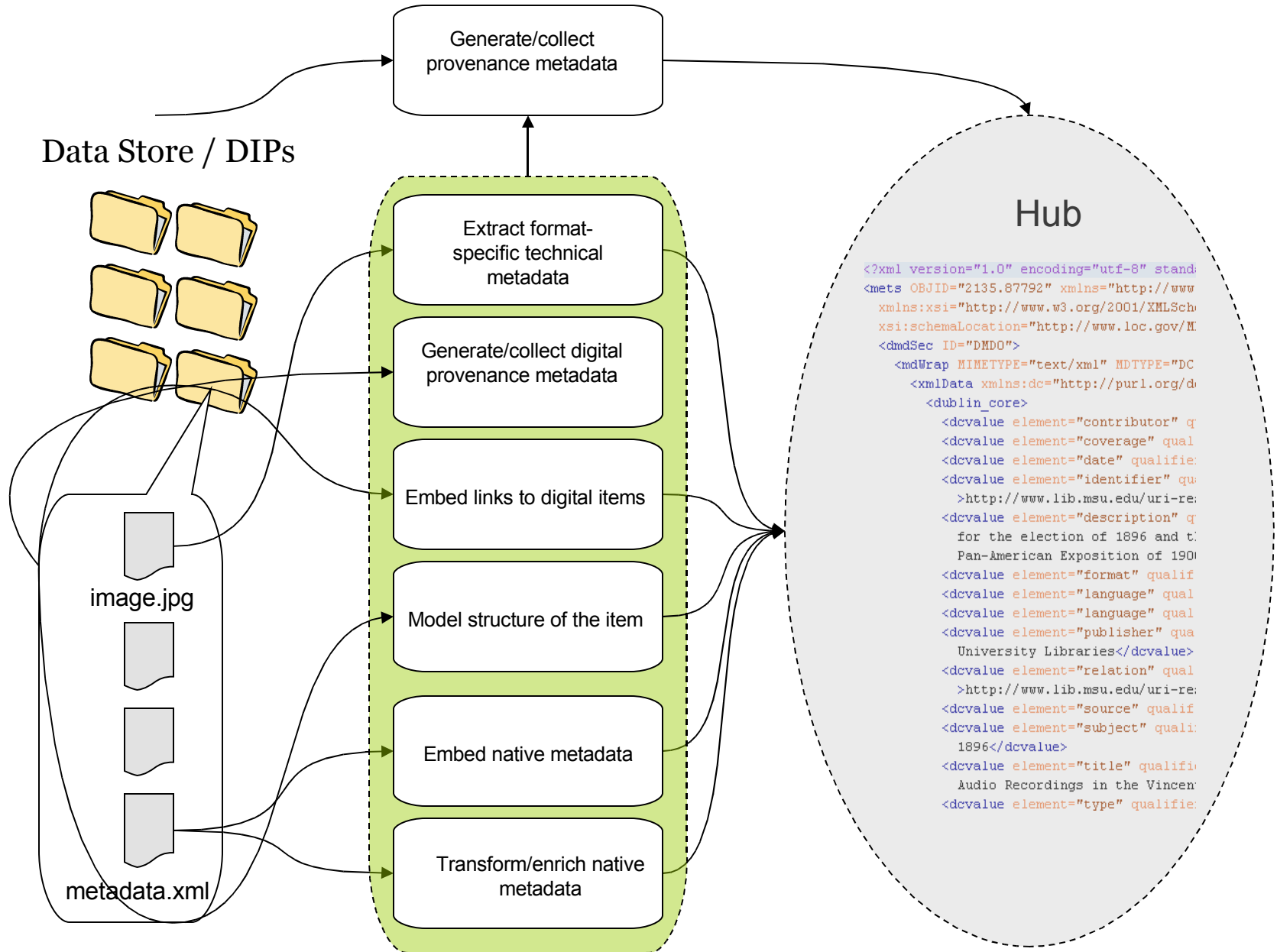
- JHOVE Output + Custom XSLT
- Java “Applicators” for specific technical metadata schemas
 - MIX
 - TEXTMD
 - AUDIOMD
 - PREMIS
 - Class hierarchy to support new Applicators



Functional Overview



To-Hub Processing



From-Hub Processing

Generate provenance metadata

Hub

```
<?xml version="1.0" encoding="utf-8" stand
<mets OBJID="2135.87792" xmlns="http://www
xmlns:xsi="http://www.w3.org/2001/XMLSch
xsi:schemaLocation="http://www.loc.gov/M
<dmdSec ID="DMDO">
  <mdWrap MIMETYPE="text/xml" MDTYPE="DC
    <xmlData xmlns:dc="http://purl.org/dc
      <dublin_core>
        <dcvalue element="contributor" q
        <dcvalue element="coverage" qual
        <dcvalue element="date" qualifie
        <dcvalue element="identifier" qu
        >http://www.lib.msu.edu/uri-re:
        <dcvalue element="description" q
        >for the election of 1896 and t.
        Pan-American Exposition of 190
        <dcvalue element="format" qualif
        <dcvalue element="language" qual
        <dcvalue element="language" qual
        <dcvalue element="publisher" qua
        University Libraries</dcvalue>
        <dcvalue element="relation" qual
        >http://www.lib.msu.edu/uri-re:
        <dcvalue element="source" qualif
        <dcvalue element="subject" quali:
        1896</dcvalue>
        <dcvalue element="title" qualifi
        Audio Recordings in the Vincen
        <dcvalue element="type" qualifie:
```

Transform hub metadata
to repository-compatible
metadata

Assemble into packages for
repository ingest

Add the METS file as
an item in the
submission package

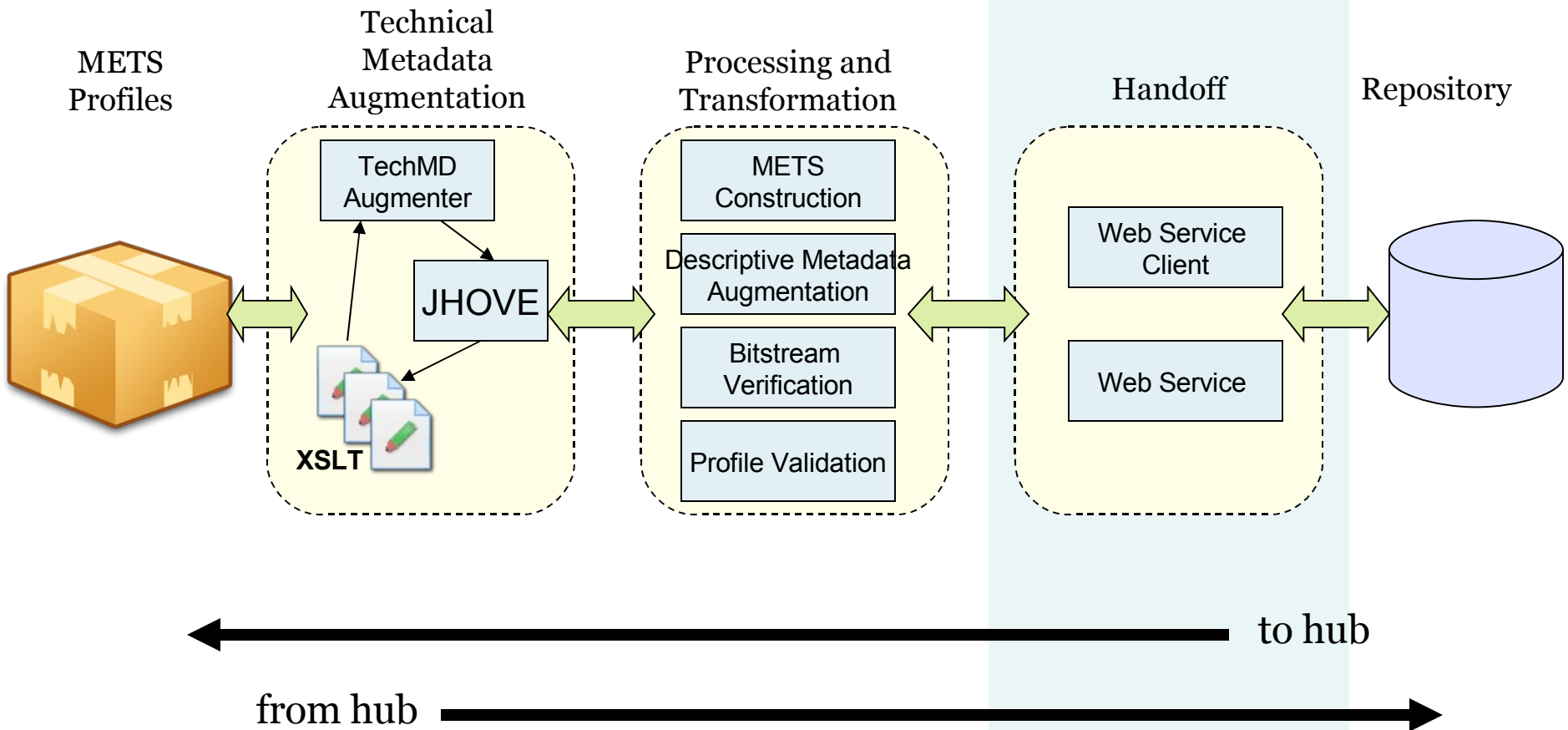
SIPs

metadata.xml

hubMets.xml



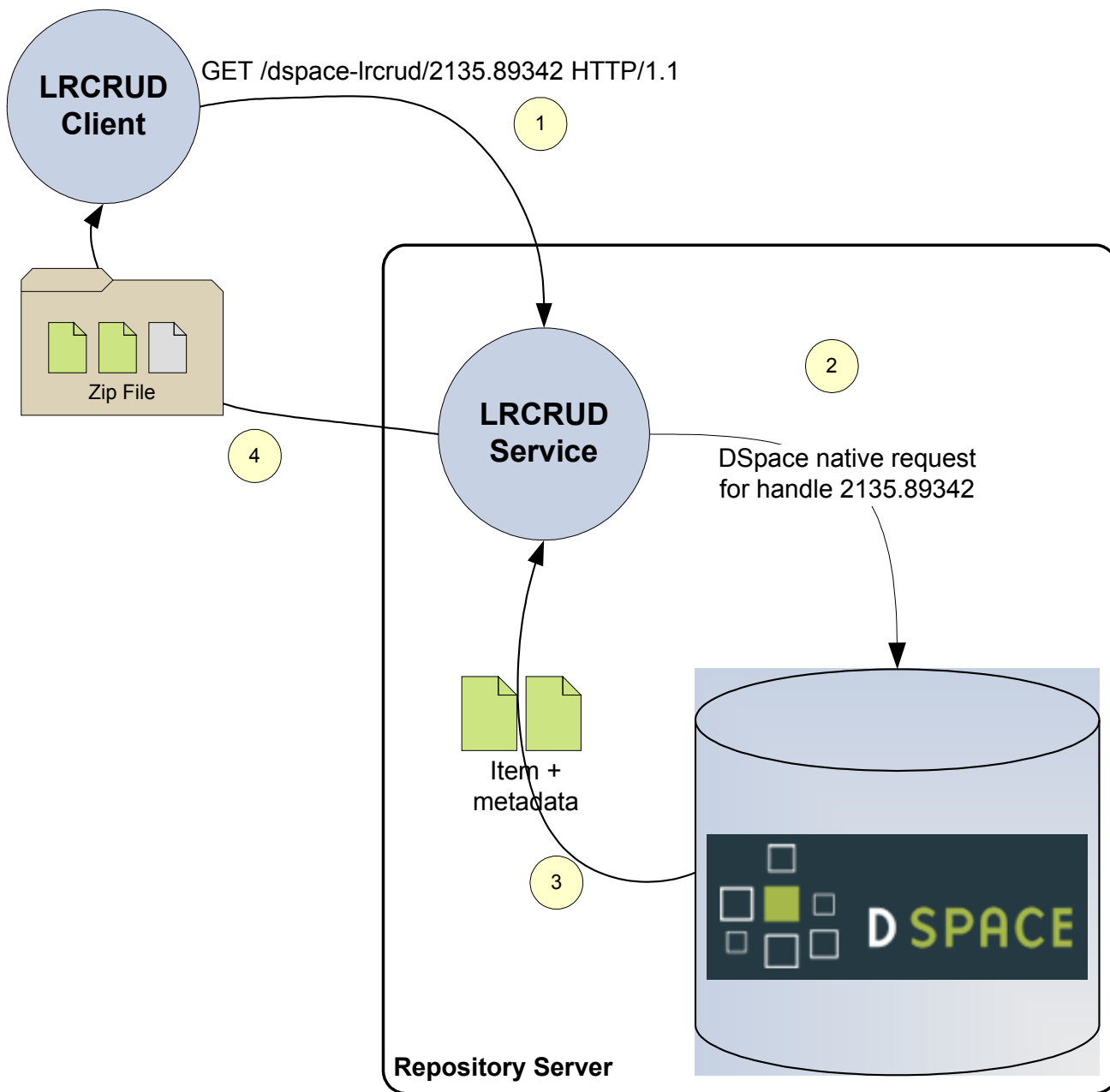
Functional Overview



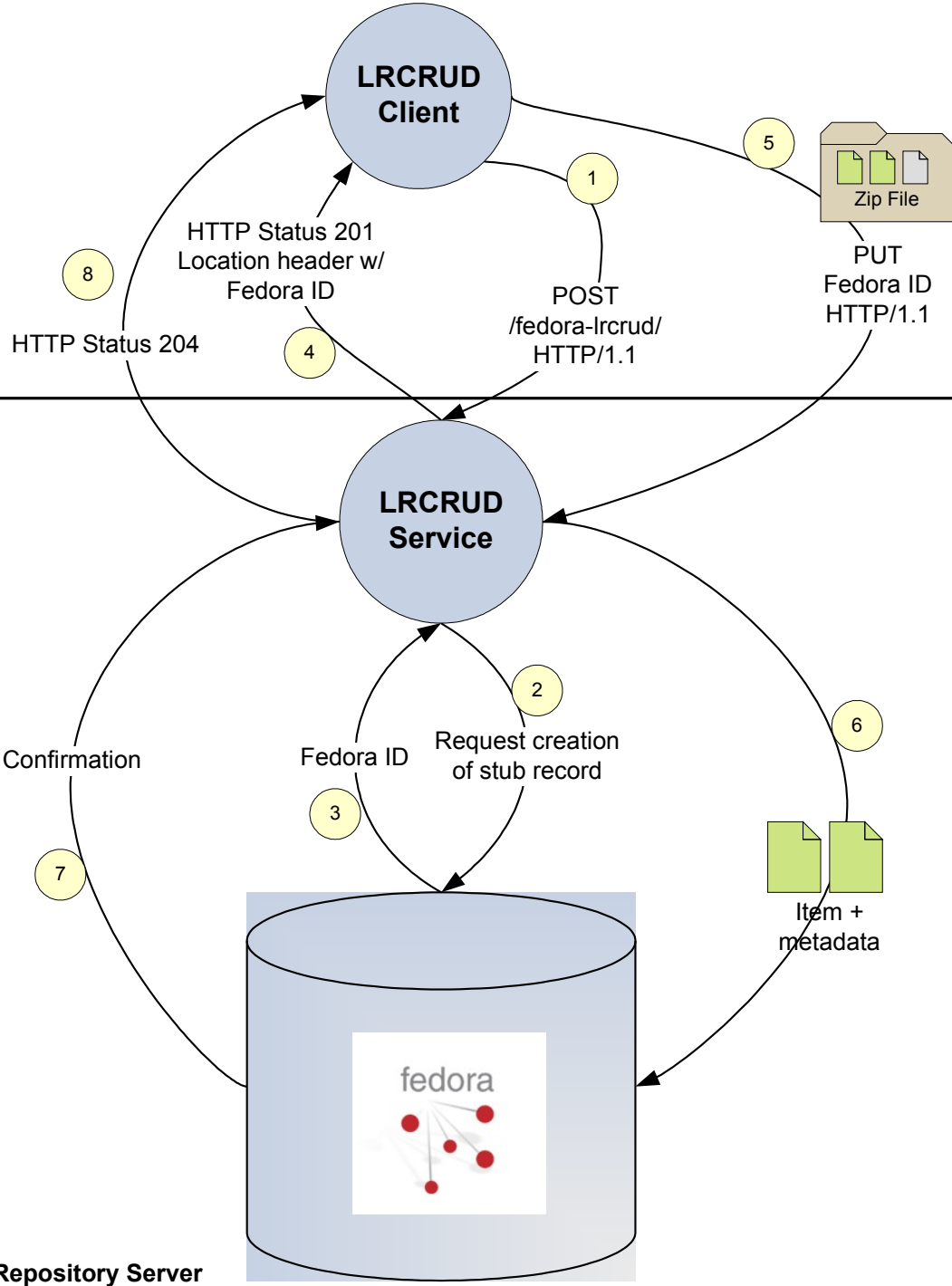


LRCRUD

- Packages usable by a repository's native ingestion routines
- REST Web-Service
 - Client integrated into processing workflow
 - DSpace, EPrints, and others in the next year
 - Specification and API to create service for other repository systems
- Similar to SWORD (Simple Web-service Offering Repository Deposit)



- Client submits a GET request to LRCRUD service for a specific item
- Service calls the native DSpace dissemination routine
- Service receives the output from the dissemination, creates a header file, and adds both the header file and the disseminated content to a zip-file
- Service returns a zip-file containing the package to the client



Create stub record

- 2) Client issues a POST request to LRCRUD specifying "where" to create the record (e.g. communities or collections) if needed
- 3) LRCRUD calls the native Fedora creation routine
- 4) Fedora supplies LRCRUD with the ID for the newly created record
- 5) LRCRUD responds to the client with an HTTP 201 "Created" message and returns the ID in the Location: header

Upload and ingest the item

- 8) Client issues a PUT request to LRCRUD to replace the package identified by the URI. The entity body of the request must contain the zip-file containing the package to be ingested.
- 9) LRCRUD unpacks the files and calls the native Fedora ingestion routine.
- 10) Fedora tells LRCRUD that ingestion was successful
- 11) LRCRUD responds to the client with an HTTP 204 "No Content" message indicating that the request was successful.







DEMO

- URLs for DEMO
 - <http://sif.grainger.uiuc.edu:8080/dspace-lrcrud>
 - <http://sif.grainger.uiuc.edu:8080/fedora-lrcrud>
 - <http://sif.grainger.uiuc.edu:8080/fedora/get/echodep:3>
 - <http://sif.grainger.uiuc.edu:8080/xmlui>
 - <http://sif.grainger.uiuc.edu:8080/sword/servicedocument>



GET

 **Hub and Spoke Workflow Manager**   

GET **PUT** **MIGRATE** **SWORD**

URL of LRCRUD Service

Export type - Item or List?

☒ **Item** ☐ **List**

Package ID or Handle (Item)

List Location

Package Export Destination

To-Hub Packager Class

 ▼

Save in Bagit format?





☐ **Yes** ☒ **No**

Log

```
INFO: Instantiating LRCRUD client
INFO: Processing item 123456789/55...
INFO: Retrieving package 123456789/55
INFO: Saving export zip file
INFO: Creating destination directory: C:\Documents and
Settings\Tom\Desktop\HandS\123456789_55
INFO: Creating Packager
INFO: Packaging files
INFO: Deleting temporary files
INFO: Done!
```



PUT

 **Hub and Spoke Workflow Manager**   

GET **PUT** **MIGRATE** **SWORD**

URL of LRCRUD Service

Collection ID or Handle

Source Type
☒ **METS file** ☐ **Zip file** ☐ **List (text file)** ☐ **Directory Crawl**

Source Location
 Browse

From-Hub Packager Class
 ▼


Keep a Local Copy of Repository Packages?
☐ **Yes** ☒ **No**
 Browse

Cancel **Submit**

Log
INFO: Found 1 items to submit
INFO: Processing
file:/C:/DOCUME~1/Tom/LOCALS~1/Temp/echodep5297017067
05734546.tmp/123456789_55/mastermets.xml...
INFO: Creating new repository package
INFO: ID = 123456789/60
INFO: Generating JHOVE metadata
INFO: Packaging files
INFO: Updating repository package
INFO: 204: No Content
INFO: ID:
http://sif.grainger.uiuc.edu:8080/dspace-lrcrud/123456789/60
INFO: Done!



MIGRATE

 Hub and Spoke Workflow Manager

GETPUTMIGRATESWORD

From

URL of LRCRUD Service

Package ID or Handle (Item) **Export type - Item or List?**

☒ Item ☐ List

List Location

Browse

To-Hub Packager Class

edu.uiuc.ndiipp.hubandspoke.packager.Dspace2HubPackag... ▼

To

URL of LRCRUD Service

Collection ID or Handle

From-Hub Packager Class

edu.uiuc.ndiipp.hubandspoke.packager.Hub2FedoraPackager ▼

Keep a Local Copy of Repository Packages?

☐ Yes ☒ No

Browse

Cancel

Submit

Log


C:\DOCUME~1\Tom\LOCALS~1\Temp\echodep529701706705734546.tmp\123456789_56
INFO: Creating destination directory:
C:\DOCUME~1\Tom\LOCALS~1\Temp\echodep529701706705734546.tmp\123456789_56
INFO: Creating Packager
INFO: Creating Packager
INFO: Packaging files
INFO: Packaging files
INFO: Deleting temporary files
INFO: Deleting temporary files
INFO: Done!

INFO: Done!

INFO: Retrieved 1 items from the repository.
INFO: Retrieved 1 items from the repository.
INFO: Processing
file:/C:/DOCUME~1/Tom/LOCALS~1/Temp/echodep529701706705734546.tmp/123456789_56/mastermets.xml...
INFO: Processing
file:/C:/DOCUME~1/Tom/LOCALS~1/Temp/echodep529701706705734546.tmp/123456789_56/mastermets.xml...
INFO: Creating new repository package
INFO: Creating new repository package
INFO: ID = echodep:8
INFO: ID = echodep:8
INFO: Generating JHOVE metadata
INFO: Generating JHOVE metadata
INFO: Packaging files
INFO: Packaging files
INFO: Updating repository package
INFO: Updating repository package



SWORD

 Hub and Spoke Workflow Manager

GETPUTMIGRATE**SWORD**

Get Service Document

Services

http://sif.grainger.uiuc.edu:8080/sword/s

DSpace 1.5.1 Hub and Spoke

New

Test

Test1

Test2

Collection Summary

Collection location	http://sif.grainger.uiuc.edu:8080/sword/deposit/123456789/2
Collection title	Test
Abstract	null

Collection Location URL

http://sif.grainger.uiuc.edu:8080/sword/deposit/123456789/2

Source Type

☒ METS file ☐ Zip file ☐ List (text file) ☐ Directory Crawl

Source Location

om\Desktop\HandS\123456789_55\mastermets.xml

Browse

Keep a Local Copy of Repository Packages?

☐ Yes ☒ No

Browse

☐ Use MD5 ☐ Corrupt MD5 ☐ Use noOp ☐ Use verbose

Slug Header

CancelSubmit

Log

INFO: Found 1 items to submit
INFO: Processing
file:/C:/DOCUME~1/Tom/LOCALS~1/Temp/echodep529701706705734546.tmp/123456789_55/mastermets.xml...
INFO: Generating JHOVE metadata
INFO: Packaging files
ERROR: Exception: null
INFO: Found 1 items to submit
INFO: Processing
file:/C:/DOCUME~1/Tom/LOCALS~1/Temp/echodep529701706705734546.tmp/123456789_55/mastermets.xml...
INFO: Generating JHOVE metadata
INFO: Packaging files
ERROR: Exception: null



More Information

Open Source Code:

<http://sourceforge.net/projects/echodep>

LRCRUD Service Specification:

<http://dli.grainger.uiuc.edu/echodep/hns/LRCRUDS.htm>

METS Profiles:

Generic -

<http://www.loc.gov/standards/mets/profiles/00000015.xml>

Web Capture -

<http://www.loc.gov/standards/mets/profiles/00000016.xml>

Java API Documentation (Javadoc):

<http://dli.grainger.uiuc.edu/echodep/hands/javadoc>

Project Web Sites

<http://ndiipp.uiuc.edu/>

<http://dli.grainger.uiuc.edu/echodep/>

Other Technologies

- Open Archives Initiative
Object Reuse and Exchange
 - <http://www.openarchives.org/ore/>



Questions?



Tom Habing
thabing@uiuc.edu